AFRL-HE-AZ-TR-2006-0015-Vol I

Air Force Research Laboratory

Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report

Brian T. Schreiber

Lumir Research Institute
195 Bluff Avenue
Grayslake IL 60030

Winston Bennett, Jr.

Air Force Research Laboratory
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

Volume I

July 2006

Final Report for March 2002 to July 2005

Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa, AZ 85212-6061

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-AZ-TR-2006-0015-Vol I

This technical report has been reviewed and is approved for publication.

// Signed //	**// Signed //**
WINSTON BENNETT JR. **Project Scientist**	**HERBERT H. BELL** **Technical Advisor**

// Signed //

DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*: 03/31/2006

2. REPORT TYPE: Final

3. DATES COVERED *(From - To)*: March 2002 to July 2005

4. TITLE AND SUBTITLE

Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report

5a. CONTRACT NUMBER: F41624-97-D-5000

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER: 62202F

5d. PROJECT NUMBER: 1123

5e. TASK NUMBER: AS

5f. WORK UNIT NUMBER: 03

6. AUTHOR(S)

Schreiber, Brian T[1]; & Bennett, Winston Jr[2]

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

1. Lumir Research Institute
195 Bluff Ave
Grayslake, IL 60030

2. Air Force Research Lab
Warfighter Readiness Rsch Div
6030 South Kent Street
Mesa AZ 85212-6061

8. PERFORMING RGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S): AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S): AFRL-HE-AZ-TR-2006-0015 – Vol I

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

This is the first volume of a five-volume report, Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study.

14. ABSTRACT

Distributed Mission Operations (DMO) training consists of multiplayer networked environments enabling warfighting training on higher-order individual and team-oriented skills. Surprisingly, only sparse DMO training effectiveness literature can be found and very few studies contain objective data. The dataset used in this research represents the largest DMO effectiveness dataset known to exist today (76 teams/384 pilots on over 3,000 engagements), containing 33 months' worth of multi-faceted DMO data, including objective data from the simulators, multiple participant surveys, subject matter expert (SME) ratings of performance, and knowledge structure tests. Observed performance differences between the pre- and post-test mirror-image point-defense assessment sessions served as the primary basis for the evaluation. Results were dramatic: On the post-test, 58.33% fewer enemy strikers reached their target and there were 54.77% fewer F-16 mortalities. Furthermore, there were corroborating significant improvements from the numerous measured skill metrics (e.g., weapons employment), SME expert observer ratings, and participant self-report opinion ratings. These converging results provide substantial evidence that pilots' become much more proficient on key aspects of combat mission objectives as a function of training within the simulator. Finding highly significant performance differences across multiple datasets between the pre- and post-tests with a combat-ready participant pool in a complex task/environment forms a formidable argument that DMO training yields considerable within-simulator warfighter competency improvement. In this report, we summarize the different dataset classes, overview the primary hypotheses and results associated with each, and discuss the convergence of the datasets to illustrate the "big picture" DMO training effectiveness. As such, more detailed hypotheses, analyses, and discussions are contained in separate reports (see AFRL-HE-AZ-TR-2006-0015, Volumes II through V).

15. SUBJECT TERMS

Distributed Mission Operations; DMO; MEC; Mission Essential Competencies; Networked environments; Skill acquisition; Training effectiveness; Warfighter training;

16. SECURITY CLASSIFICATION OF: Unclassified

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLASSIFIED	UNCLASSIFIED	UNCLAS	UNLIMITED	37	Dr. Winston Bennett Jr
					19b. TELEPHONE NUMBER *(include area code)*

This page intentionally left blank

Table of Contents

List of Figures and Tables

EXECUTIVE SUMMARY

Distributed Mission Operations (DMO) training consists of multiplayer networked environments enabling warfighter training on higher-order individual and team-oriented skills—areas identified as training "gaps" by operational pilots. Surprisingly, convincing DMO training effectiveness studies are lacking. This research examines the largest DMO effectiveness dataset known to exist (384 pilots on over 3,000 engagements containing over 22,000 threats and over 35,000 munitions employed). Over 55 billion individual data points were collected from the simulators, over 1,400 subject matter expert observer evaluation sheets were completed, 1,728 participant surveys were administered, and all 384 pilots were asked to complete a Pathfinder knowledge structure task. The objective was to report a large-scale, scientifically-sound, comprehensive within-simulator DMO training effectiveness baseline evaluation with different, but complimentary datasets expected to converge on similar conclusions regarding the overall learning benefit of DMO. In this report, we summarize the four dataset classes, overview only the primary hypotheses and results, and discuss the convergence of the datasets to illustrate the "big picture" DMO training effectiveness. As such, more detailed hypotheses, analyses, and discussions are discussed in AFRL-HE-AZ-TR-2006-0015, Volumes II through V).

Seventy-six F-16 teams participated in five days of DMO training research. Observed performance differences between the pre- and post-test mirror-image point-defense benchmark assessment sessions served as the basis for the evaluation. Results were quite dramatic: On the post-test, the teams, on average per scenario, allowed 58.33% fewer enemy strikers to target, killed 9.20% more enemy aircraft, permitted 54.77% fewer F-16 mortalities, spent 55.20% less time allowing hostiles into MAR, and 60.33% less time allowing hostiles into N-pole ranges ($p < .01$ for all). Expert observer ratings—both those taken in real-time and those done according to a scientifically blind protocol—revealed statistically significant improvements as a function of DMO training. These improvements were found both for briefs/debriefs and also for mission execution, corroborating the objective results. Surveys of participating pilots and Airborne Warning and Control System (AWACS) operators showed a strong acceptance of DMO as a training device. "I would recommend this training experience to other pilots/controllers" was rated by all but one of 49 controllers and all but 16 of 327 pilots with the highest rating possible of "Strongly Agree." The pilots also rated the Mesa DMO environment higher than all seven other training environments surveyed for providing training utility on the Mission Essential Competency (MEC) experiences.

The results reported here provide very strong evidence that pilots become more competent in the simulator as a function of DMO training. There were a number of factors in this research that we anticipated would undermine the chances of revealing statistically significant within-simulator DMO learning effects, including:

(a) an applied research study on an extremely complex and ecologically valid task,

(b) changes to the experimental environment creating additional noise variance in the data,

(c) missions/tasks that can be only partially controlled, and

(d) a highly experienced, combat-ready participant pool whose performance levels arguably may have been at or approaching asymptote before ever participating in the current effort.

Finding highly significant performance differences between the pre- and post-tests across many different data sources in light of these factors provides a formidable argument that DMO training yields considerable within-simulator warfighter competency improvement. The successive volumes to this report (Volumes II through V) contain more detailed procedures, results, and discussion points.

ACKNOWLEDGEMENTS

The opinions expressed within this report are those of the authors and do not necessarily reflect the views of the sponsoring/employing organizations. This work was funded by the U.S. Air Force under contract #F41624-97-D-5000. Funding for the effort was jointly provided by the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division (AFRL/HEA) and the Air Combat Command (ACC). Because this research took many years to complete and required the assistance of much of the AFRL/HEA organization, we simply could not list all significant contributors as co-authors. The authors wish to extend a very special thank you to the following individuals (in alphabetical order) for their significant contribution to the current work, the vast majority of whom put in hundreds of hours supporting this effort. Without them, this report could not have been possible:

Dr. George Alliger: Research Psychologist
Michelle Borlase: Data Collector
Lt Col. Barry Brannon: SME, Data Collector
Lance Call: DMO Engineer
Lt Col. Jeff Carr (ret.): Participant Liaison
Lt Col. Chuck "DBear" Colegrove (ret.): Air Combat Command Liaison
Jason Cox: DMO Engineer
Col Todd Denning (ret.): SME
Pam DiSalvo: Researcher & Volume V co-author
Tim Esplin: DMO Engineer
Mike Forer: Database Specialist
Lt Col. Mike "Frenchy" France (ret.): SME, Data Collector
Dr. Sara (Liz) Gehr: Researcher & Volume III co-author
Lt Col. Rick Glitz: SME, Data Collector
Dave Greschke: DMO Manager
James Hartman: Data Collector
Mary Johnson: Technical Writer
Maj. Kenneth Kleinlein: SME
Lt Col. Randel "Cow" Lane (ret.): SME, Data Collector
Jim Marx: Console Operator
Lt Col. Chris "New York" Metz (ret.): SME, Data Collector
Lt Col. Mike "Odie" Park (ret.): SME, Data Collector
Maj Les "Mo" Pogue (ret.): SME, Data Collector
Antoinette Portrey: Human Factors Engineer
Lt Col. Bart Raspotnik (ret): SME, Mission Director, DMO Engineering Team Lead
Maj Robert "Bobaloo" Rickard: SME, Data Collector
Brett Rigsby: Software Engineer
Lt Col. Robbie Robbins (ret.): SME, Mission Director
Capt Ira Schurig: Behavioral Scientist
Margaret Schurig: Researcher & Data Collector
Leah Rowe: Researcher & Volume IV co-author
Traci Smith: Research Assistant, Data Collector
Don Smoot: Mission Director
Dr. William A. Stock: Researcher & Volume II/V co-author
Scott Swigert: DMO Engineer
Lt Col. Steve Symons: SME, Syllabus Developer, Data Collector
Lt Col. Mark "Pop" Warner (ret.): SME, Data Collector
Eric Watz: Software Engineer
Mike Williamson: DMO Engineer
Lt Col. Jack Wooster (ret.): SME, Mission Director

DISTRIBUTED MISSION OPERATIONS WITHIN-SIMULATOR TRAINING EFFECTIVENESS BASELINE STUDY: SUMMARY REPORT, VOLUME I

INTRODUCTION

General Problem

A paradigm shift is occurring within the United States Air Force (USAF) today. The Air Force is augmenting its frequency-based training system, known as the Ready Aircrew Program, or RAP (United States Dept. of the Air Force: Flying Operations, 2002) with a competency-based training system. A promising recent process methodology identifies the Mission Essential Competencies (MECs) necessary for an individual, team, or crew to be successful in combat under adverse conditions. The MEC process, driven by data from operational warfighters, identifies the skills necessary for combat and the experiences required to become proficient in those skills (Colegrove & Alliger, 2002). Competency-based training defines a standard level of *proficiency* or *competency* in each skill to be achieved, philosophically and functionally quite different from a frequency-based program that mandates a given number of various types of missions to be performed. Due to this fundamental difference, competency-based training assessment emphasizes training the skill, knowledge, and experience deficiencies (or “gaps”) for individuals, teams, or crews. Distributed Mission Operations (DMO) training, especially those using networked simulators, is often mentioned as a viable training medium for fulfilling many skill and experiential deficiencies.

For air combat — the domain under study in the current work — many of the competency gaps revolve around higher order tasks or skills that can be gained from more complex experiences (e.g., team work, multi-team operations, complex tactical maneuvers, etc.). In early work unrelated to the MEC process but yielding some relevant results, researchers surveyed 94 F-15 air combat pilots and also discovered higher order experiential areas as receiving less than adequate training in their current unit, specifically: Multi-bogey, reaction to surface-to-air missiles (SAMs), dissimilar air combat tactics, all-weather employment, electronic countermeasures/electronic counter countermeasure (ECM/ECCM) employment, communications jamming, low altitude tactics, chaff/flare employment, escort tactics, threat early warning system (TEWS) assessment, and work with the Air Weapons Controller (Houck, Thomas, & Bell, 1991). For each of the aforementioned training areas, the F-15 pilots also felt those training areas were also better suited to the simulator--precursor opinion-based evidence for the potential of emerging DMO environments. Gray, Edwards, and Andrews (1993) interviewed 99 F-16 pilots, asking a number of open-ended questions. The subsequent content analyses revealed that the highest reported “difficult aspects” of attaining/maintaining mission-ready status were weapons delivery, radar interpretation, electronic combat, cockpit switchology, and air-to-air combat. Roughly two-thirds of the pilots perceived a significant loss of knowledge/skill between completing schoolhouse training and entry into the operational unit. Again an early indication of DMO’s perceived potential, when the F-16 pilots were asked which ground-based media they would like to see used more, the most preferred option was the simulator. More recently, as part of the MEC process, operational warfighters identified the experiences that contribute to the development of becoming a successful warfighter in combat. As part of ongoing MEC data collection, operational warfighters have been surveyed as to the utility of DMO in providing those training experiences. Depending on DMO site, operational F-15 and F-16 pilots have consistently reported at least half (and even as high as over three fourths) of their critical MEC experiences could be gained “to a moderate extent” or better in DMO.

In stark contrast to stand-alone simulators of the past that primarily served to train emergency procedures or other routine tasks, DMO training consists of multiplayer networked environments enabling frequent warfighting training on higher order individual and team-oriented skills. Most DMO training environments consist of multiplayer high-fidelity networked simulators. These networked environments allow geographically distributed warfighters (local and/or long-haul) the ability to come together as a team or crew and train against manned and/or simulated adversaries (Callander, 1999; Chapin, 2004). This DMO training therefore affords opportunities to gain battle-like *experiences* not frequently gained outside of war. Examples of DMO simulation environments include the F-16 mission training center at Shaw Air Force Base, the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division’s (AFRL/HEA’s) DMO training research site in Mesa, AZ, the F-15 mission training centers at Langley and Eglin, the multi-ship Jaguar simulation facility in Bedford, England, etc. DMO training capabilities can be generally defined as affording the ability to bring a number of warfighters together to train complex individual and/or team tasks during the course of larger scale, realistic combat missions (Chapman, Colegrove, & Greschke, in press).

DMO simulation training environments are relatively new. Until the late 1990s, warfighters received training on complex tactical missions almost exclusively during infrequent larger scale range exercises. Because of this, we speculate that a survey of operational pilots 20 years ago would have revealed far fewer pilots identifying DMO as a viable training gap filler; it is likely that results would have revealed more frequent range exercises as a more favored solution. However, even realistic range exercises posed then and continue to pose today many training restrictions. At Red Flag, a USAF large-scale range exercise (Boyne, 2000), there are space and altitude restrictions governing all aircraft. In addition to being costly, resource availability limits the potential number of aircraft. The maneuver restrictions and the limited frequency of range exercises constrain the warfighter’s training compared to how he/she would actually fight in war. With the advent of DMO training, resource issues and tactical employments are significantly less restricted, thereby allowing warfighters more opportunities to train for wartime requirements of today or possibly even to train to the potential wartime requirements of tomorrow. These logistical advantages combined with the previously discussed survey research suggest that DMO appears to be one promising environment for competency-based training. But, just how much more competent do our warfighters become as a function of training in DMO environments? That is, just how effective is DMO training?

Training Effectiveness Evaluations: Literature Review

Convincing effectiveness evaluations consist of different types of data converging on the same conclusions. Kirkpatrick (1975) provided four levels for evaluating training--trainee perceptions, measured evaluations of learning, observed performance, and impact. Another well-known evaluation model, Bell and Waag (1998) offered a similar framework, one where data is collected from warfighter opinions, instructor or expert rater observations, and objective data collected both in the simulator and on comparable “real-world” transfer tasks. Common to both these evaluation models, a proper and thorough evaluation necessitates multiple sources of data that (ideally) converge. Central to the evaluation, objective data enables quantifying the training effectiveness by measuring improvements in mission outcomes and skill proficiency, thereby providing indications of the return on investment (ROI), in terms of increased human performance, of the training system. Instructor or rater observation data provides expert assessment of skill competency, corroborating the objective data. Rounding out the evaluation, user opinion data captures what the users experienced and their opinions on the usefulness of the training system, its pros and cons, and which tasks the system might be best suited for.

Using various degrees of opinion, rater, and/or objective data, a fair amount of prior simulator training effectiveness research exists for *simpler* tasks representative of a small portion of a mission (e.g., manual bomb delivery, one versus one air combat) and all found simulator training beneficial (e.g., Gray, Chun, Warner, & Eubanks, 1981; Gray & Fuller, 1977; Hagin, Dural, & Prophet, 1979; Hughes, Graham, Brooks, Sheen, & Dickens, 1982; Jenkins, 1982; Kellogg, Prather, & Castore, 1980; Leeds, Raspotnik, & Gular, 1990; Lintern, Sheppard, Parker, Yates,& Nolan, 1989; McGuinness, Bouwman, & Puig, 1982; Payne, et al., 1976; Robinson, Eubanks,& Eddowes, 1981; Wiekhorst & Killion, 1986; for reviews, we also refer the reader to Bell & Waag [1998] and Waag [1991]). Compared to predominantly stand-alone systems of the past, DMO training not only affords the ability to train team skills, but also to train larger, more complex portions of the mission and higher order individual cognitive skills. Given that contemporary DMO environments afford the ability to train very different and more varied skills, at best only limited generalizations can be drawn from the above-cited historical training effectiveness research.

Some more recent and relevant multiplayer simulation research suggests DMO enhances individual and team skills for:

- F-15 pilots (Houck, Thomas, & Bell, 1991),
- F-16 pilots (Berger & Crane, 1993),
- Tornado pilots and navigators (Huddlestone, Harris, & Tinsworth, 1999), and
- pilots, forward air controllers, and ground forces executing close air support (Bell, et al., 1996).

F-16 pilots who have flown in a distributed environment have rated DMO as a particularly effective training system for missions involving 4-ship air-to-air employment against multiple enemy aircraft (Crane, Schiflett, & Oser, 2000). F-16 pilots have reported that both individual skills (such as radar mechanization, communication, and building situation awareness) and team skills (such as maintaining mutual support, tactical execution, and flight leadership) are enhanced by DMO training (Crane et al., 2000). Many of these studies have heavily relied on subjective assessments, an assessment method for DMO that we have discovered to be useful, but one that still possesses assessment issues. These include potential vested interest and bias by raters, lack of measurement sensitivity, and an inability to correctly track simple statistics such as kills (Krusmark, Schreiber, & Bennett, 2004). Ideally, objective data would provide the DMO assessment foundation, with multiple sources of augmenting data (e.g., expert ratings, participant surveys) to complement and to converge on the effectiveness conclusions.

Over the past several years, attention and resources have been focused on DMO training and supporting technologies, the focus of which has generally been on engineering improvements to create a more realistic environment. Future enhancement efforts have generally revolved around addressing questions such as "What" in the simulation environment is not realistic and "How" can we make it more realistic (Watz, Schreiber, Keck, McCall, & Bennett, 2003). Amazingly, all this improvement effort comes without a documented training effectiveness baseline for DMO, let alone a scientific understanding of which technologies should be pursued first. Of course, DMO training environments exist first and foremost to <u>improve warfighter competence</u>, not necessarily to create the most realistic environment as an end unto itself. Investigating and evaluating warfighter competency as a function of DMO training invites addressing entirely different, non-engineering questions, such as how much better do our warfighters become from DMO training, which MEC skills are best trained in DMO, and quantitatively, what are the improvements? Though very few doubt DMO training as beneficial, the literature does not provide irrefutable evidence as to the magnitude and types of human performance gains. The literature trends towards improvement, but the evidence is relatively sparse for many vs. many DMO environments and is based almost solely upon subjective data. Indeed, the disturbingly low degree of documented effectiveness evidence is not new (Waag, 1991). A comprehensive DMO training effectiveness evaluation would serve to help justify the expenditures on these environments, to provide an effectiveness baseline study to evaluate against when investigating changes to a DMO environment, and to provide an initial assessment as to which competencies are best trained in a DMO environment.

To fully understand the true benefits and potential of DMO, a number of studies need undertaking. One of the first studies needs to serve as a baseline; a study which documents and quantifies the amount of learning occurring as a function of training time spent in the DMO environment. That is, just how much more competent are our warfighters as a function of training in these environments? Since DMO training today is largely network simulation-based, we will refer to this baseline learning effect as the degree of "within-simulator" learning. Once within-simulator learning has been established and quantified as a baseline, subsequent studies would investigate the logical, follow-up robustness and application questions, such as how quickly does the within-simulator learning effect decay, how much of the learning effect transfers to the "real world", how many and which skills are best trained in the DMO environment, which technological enhancements provide the best return on increases in human performance, etc. Very few studies exist addressing these scientific questions for DMO in general, and its more specific impact on competency-based approach to training and assessment. Therefore, the current work sought to provide a scientifically proper and quantifiable DMO within-simulator learning training effectiveness baseline—a potential landmark study to be used for calculating warfighter performance return on investment in DMO and a study used as reference for other, future DMO robustness and application studies.

CURRENT WORK

Some of our "early look" effectiveness results have been previously published and provide strong initial indications of DMO within-simulator training effectiveness (Gehr, Schreiber, & Bennett, 2004; Schreiber, Watz, Bennett, & Portrey, 2003). Similar to the current work, each of those studies examined F-16 pilots before and after five days of DMO training, reporting substantial improvements in mission outcomes. The preliminary studies, combined with the prior (but sparse) DMO effectiveness literature, suggest that DMO training provides an extremely effective environment for improving air combat competencies. This research builds upon the work from both Schreiber et al.(2003) and Gehr et al.(2004) by examining the largest DMO effectiveness dataset known to exist. This effort, reported in a series of five volumes, aims to report substantially more data and more thorough analyses than our first-look studies. The overall objective of this volume is to summarize the reporting of a large-scale, scientifically sound, comprehensive within-simulator DMO training effectiveness baseline evaluation, with objective data, subject matter expert (SME) observer rating data, pilot self-report opinion data, and knowledge structure data.

GENERAL METHOD

After F-16 pilots arrived at the AFRL/HEA DMO training research facility in Mesa, AZ, they received some simulator familiarization training and then were immediately "benchmarked," or "tested" on their pre-training point defense scenario performance. Post-training reassessment with those same pilots using mirror-image point defense scenario benchmarks occurred at the completion of five-day DMO training. Observed performance differences on 76 teams between the pre- and post-test benchmark assessment sessions served as the basis for the within-simulator training effectiveness evaluation. We collected a variety of DMO effectiveness data from numerous sources and organized them into four major dataset classes. In this report, we summarize each dataset class, overview the primary hypotheses, report the high-level results, and discuss the convergence of the datasets to illustrate the "big picture" within-simulator DMO training effectiveness. More detailed hypotheses, methods, procedures, results, and discussion for the different dataset classes are reserved for separate, more detailed stand-alone reports (see Volumes II through V).

Indices of DMO training effectiveness

Dataset Class A includes the objective outcome measures and process/skill measure databases collected during the Monday and Friday pre- and post-test benchmarks. Billions of individual data points from over 3,000 engagements were collected and aggregated. The aggregated objective data collected from the mirror-image pre-/post-test scenarios serves as the cornerstone dataset for establishing DMO within-simulator training effectiveness. Outcome measures reported include enemy strikers reaching base, closest distance achieved by strikers, F-16 mortalities, and enemy striker and fighter mortalities. Process/skill and supporting competency measures reported here include weapons employment metrics, weapons engagement zone management metrics, wingman formation metrics, and communication use. We refer the reader to Schreiber, Stock, and Bennett (2006b) for more extensive objective data metrics, results, and analyses.

Dataset Class B includes all SME observer rating data collected during the pre- and post-test benchmarks. Two SME rating datasets were collected—SME ratings provided in "real-time" while pilots were flying their missions and blind ratings (ratings done at a later date using recorded benchmarks without any SME knowledge of team or pre-/post-test benchmark condition). Over 1,400 gradesheets were completed by SMEs in real-time during the full training week and another 153 gradesheets of just benchmarks were completed later using the scientific blind protocol. For this summary report, Dataset Class B serves primarily to use the expert rating data as validation/corroboration of the findings from Dataset Class A. We refer the reader to Schreiber, Gehr, and Bennett (2006c) for additional hypotheses and detailed analyses.

Dataset Class C contains all the participant opinion data collected via surveys during the familiarization session and/or at the end of the training week. A total of 1,728 surveys were administered, including demographics, DMO Feedback forms, DMO Reaction ratings, and ratings of to what extent MEC experiences could be gained in various training environments. Dataset Class C serves to report the user acceptance of DMO training and its perceived utility level and effectiveness. For this summary report, we overview the operators' opinions of DMO as a training system. We refer the reader to Schreiber, Rowe, and Bennett (2006) for more detailed analyses.

Dataset Class D includes F-16 Pathfinder data collected just before the familiarization session and again after the last training session (either before or after the post-test benchmarks). Pathfinder measures changes in knowledge structures and was used in this study to ascertain if the pilots had significant changes in their air combat knowledge structures as a function of the DMO training. We refer the reader to Schreiber, DiSalvo, Stock, and Bennett (2006) for detailed analyses.

Hypotheses

Primary hypotheses are summarized below; refer to Volumes II-V for secondary hypotheses.

1. We hypothesize that *highly* significant improvements in the Monday to Friday benchmark comparison will be observed for a number of objective indices, both in outcome-oriented metrics and in process-oriented skill metrics.
2. We hypothesize that we will not observe a significant trade-off in the observed Monday to Friday performance. That is, pilots will demonstrate improved performance on both offensive *and* defensive skill-related measures.
3. We hypothesize that significant Monday to Friday benchmark improvements will also be observed in the SME observer rating data (both real-time and blind), corroborating the objective results.
4. We hypothesize that analysis of the pilot surveys and rating forms will show DMO user acceptance, perceived utility of DMO training, and self-reported learning as a function of DMO training.
5. We hypothesize that there will be a change in pilots' knowledge structures, specifically, (a) knowledge structures will increase in similarity to expert knowledge structures, and (b) knowledge structures will have less variability between team members.

Participants

From January 1, 2002 to October 22, 2004, 76 fighter pilot teams participated in the current DMO within-simulator training research study at the Mesa DMO site. An estimated 20% of the USAF F-16 *population* -- 384 pilots -- participated in this study. To participate in the training research, operational F-16 squadrons vied for posted vacant DMO training research weeks at the Mesa research site, readily volunteering for available training research opportunities. Therefore, participants in this study were not randomly sampled. Of the 76 teams and 384 pilots under investigation, the following number of participants produced useable data for Dataset Classes A, B, C, and D:

Dataset Class A: The 53 teams (272 pilots) produced data useable for objective analyses, and all but three were male, with a mean age of 33.1 years, 10.8 average years of military service, and a mean number of hours in an F-16 of 1,016.

Dataset Class B: For the current work, we used a legacy gradesheet. Frequent testing of a new, alternative subjective assessment tool in development prevented some data collection with this gradesheet; useable SME rating data were still collected for a sizeable sample of 148 pilot participants from 37 teams--146 male and 2 female with a mean age of 32.8, 10.4 years of military service, and a mean number of hours in an F-16 of 905.7.

Dataset Class C: Several surveys were administered, and as many as 327 pilots and 49 AWACS produced useable data for one or more surveys. All but two of the 327 pilots were male, with an age range between 24 and 54 years (mean = 33.0). The pilots averaged 1,681 flight hours up to the time they participated in DMO at Mesa, and an average of 1,039 of the 1,681 total hours were F-16 hours. AWACS demographic information was available for 45 of the 49 participants. All but three of those 45 controllers were male, with an age range between 24 and 41 years (mean = 30.4).

Dataset Class D: Overall, 144 pilots were included in the Pathfinder analyses. Of these, all but two were male, with an average age of 32.3 years, an average of 9.9 years of service, and average number of 986.4 hours in an F-16. A large percentage (38.8%) of the data was eliminated for failure to meet coherence (.20), an analytical technical criterion.

The differential sample sizes for each Dataset were then used in subsequent analyses for each respective Dataset Class. This approach was favored over dataset standardization in order to maximize each dataset's sample size.

DMO Training Facility

In conjunction with a computer-generated threat system and an instructor operator station (IOS), the DMO research environment in Mesa, AZ consisted of four high-fidelity F-16 simulators and one high-fidelity Airborne Warning and Control System (AWACS) simulator. The F-16s, AWACS, and threat entities interoperated according to Distributed Interactive Simulation (DIS) standards (IEEE Standard for Distributed Interactive Simulation - Application Protocols, 1995) version 4.02 or version 6.0.

The high-fidelity F-16 Block 30 simulators utilized 360 degree out-the-window visual displays with either SGI Onyx II Reality Monsters or PC Nova IIs running Aechelon runtime software. The visual system used high resolution photo-realistic databases of the Sonoran desert overlaid on terrain elevation data of the region. The hardware was very nearly identical to that found in the actual F-16, as was the software (Software Capabilities Upgrade version 4). Depending on the type of mission to be flown, F-16 weapon load-outs for missions consisted of differing combinations of the gun, the Air Intercept Missile (AIM-9), the Advanced Medium Range Air-to-Air Missile (AMRAAM), and/or the Mk-82 and Mk-84 general purpose bombs. A high-fidelity Solipsys version 6 AWACS sensor simulation was also used to provide a more realistic environment.

The Automated Threat Engagement System (ATES) generated all adversaries. A computerized, real-time threat generation system, ATES operates on standard DIS networks, providing air-to-air, air-to-ground, and surface-to-air threats. The ATES incorporates aerodynamic modeling, atmospheric models, radar models, infra-red models, and data parameter tables for thrust, drag, lift, etc. For the current work, threat air models were the MiG-29, MiG-27/23, and Su-27 loaded with the AA-8, AA-10a, and AA-10c air-to-air missiles. Ground threats included the SA-2, SA-6, and SA-8, and antiaircraft artillery (AAA). Threat aircraft performed maneuvers and/or scripted flight paths while reacting to the F-16's maneuvers and weapons.

The debrief facility included five 50-inch plasma screens -- one for a God's eye view and one dedicated for each of the four F-16s. Each of the F-16 plasma screens presented four avionic displays from the F-16. The time synchronized replay included all communications and could be paused, fast-forwarded, or rewound according to the lead pilot's desired use of the allotted debrief time. This debrief facility was also used for the SME blind ratings of recorded missions.

As a training research installation striving to continually integrate and evaluate new training technologies, the DMO site at Mesa undergoes occasional upgrades to its simulation systems. Therefore, the DMO simulation environment was *not* constant for all participants in this study. Some examples of upgrades/changes to the environment during the 33-month data collection period included (but is not limited to):

- Upgrading the visual databases in cockpits #3 and #4 to use the same photospecific database used in cockpits #1 and #2, upgrading to eight visual channels,
- upgrading the radios,
- installing SCU-5 SADL (Situation Awareness DataLink) software,
- installing new ALQ-213 radar warning/electronic counter measure panels and 5100 power PC boards,
- adding smoke trails to missile fly-outs,
- upgrading the brief/debrief facility with Portable Flight Planning Software version 3.2, and
- a sixth 50-inch plasma debrief display for AWACS.

Under most circumstances changing the apparatus during the course of a scientific study threatens the study's conclusions. However, for the current work, we viewed these changes in the DMO environment as highly desirable. Further explained, as a system of integrated technologies, *all DMO environments will change* and be constantly upgraded at every field location. By doing similarly in our experimental environment we more closely replicate the actual systems to which we aim to generalize. Furthermore, we argue that significant learning effects must be found in light of the additional error variance associated with updates/changes to the environment, because the DMO environments will undoubtedly undergo change. If a training effect is not found under these changing conditions, justification for DMO training does not exist.

Training Research Syllabi/Training Research Week

Table 1 shows a general timeline for each participating team. Participants arrived early Monday morning for five days of DMO participation. Upon arrival, participants were first given an inbrief on the objectives and procedures of DMO and the simulators, a tour of the facilities, and then given a research administrative session where they completed a demographic form, were assigned anonymous barcode identification numbers, and finally took the first Pathfinder exercise-- an electronic assessment used to build mental models of novice and expert pilots.

Table 1 Participant General Timeline.

Session#	1	2	3	4	5	6	7	8	9
Day/time	Mon AM	Mon PM	Tues AM	Tues PM	Wed AM	Wed PM	Thur AM	Thur PM	Fri AM
Activity	Mesa Inbrief Admin Pathfinder Pilot Brief Fly Fam Pilot Debrief	Pilot Brief Fly 3 Benchs+ Pilot Debrief Feedback Survey	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 4-8 engmnts Pilot Debrief	Pilot Brief Fly 3 Benchs+ Pilot Debrief Feedback Survey Reaction Survey Pathfinder Outbrief

Pilots participated in one of four very similar syllabi, each syllabus consisting of nine 3.5 hour sessions, beginning with session one on Monday morning and ending with session nine on Friday morning. There were two sessions each day of the five-day training week, save Friday's single session. Each session entailed a one-hour briefing, an hour of flying multiple engagements of the same mission genre, and an hour and a half debriefing. The syllabi scenarios could be either offensive or defensive, but were all four F-16s versus X number of threats. Scenarios were designed with trigger events and situations to specifically train MEC skills (Symons, France, Bell, & Bennett, 2006). These syllabi were developed with traditional building block methods using full mission rehearsal scenarios across a spectrum of probable air-to-air missions and threats while increasing the complexity of the missions as the training research week progressed.

After completing the administrative tasks early Monday morning, each syllabus began with a familiarization session (session one) late Monday morning to orient pilots to DMO simulator environment specifics, such as visual ID characteristics and any switchology differences due to F-16 block number or F-16 mission software. The pilots required surprisingly little familiarity training. The hour allotted turned out to be more than enough familiarity time, as the high fidelity simulator layout and underlying simulation models closely resembled the actual aircraft and pilots quickly became comfortable with DMO simulator operation. Since the pilots readily and easily adapted to the simulation environment during the familiarization period, performance increases observed throughout the course of the subsequent sessions should be the result of learning/honing their skills and not learning "sim-isms" or other DMO idiosyncrasies.

Session two on Monday afternoon began with benchmarks (i.e., a "pre-test") used to measure pre-training performance. The training week ended with the "post-test" training benchmark session nine on Friday morning. The benchmark sessions consisted of flying 3-point defense engagements (examples are provided in Figure 1). All benchmark point defense scenarios pitted the four participant F-16s and their AWACS controller against eight threats (six hostiles and two strikers) at a distance greater than 40 nautical miles. During all benchmark scenarios, AWACS informed the F-16s (at long range to the threats) that there were six entities and that all six were already identified as hostile, thereby allowing the F-16s to shoot beyond visual range at those six entities. Regarding the two strikers, the AWACS operator could not "see" below 10,000 feet--the altitude under which the enemy strikers flew during all benchmarks. Therefore, the onus fell upon the F-16s to find any entities below 10,000 feet with their onboard radars and visually identify them before employing ordnance.

All benchmarks were designed to be equally complex according to a complexity scoring scheme outlined by Denning, Bennett, and Crane (2002). Seven-point defense benchmark scenarios were developed, and the complexity analysis revealed that all benchmarks were indeed equally complex. Pilots flew in the same flight/cockpit assignment on Monday and Friday. Unbeknown to the pilots, for the Friday benchmarks, pilots flew the mirror-image of the three benchmarks that were flown on Monday.   Strict data collection protocol governed all benchmarks in order to maintain a realistic combat environment—i.e., no freezing or reloading entities, fuel always on, no reincarnating entities, no inserting new entities, real-time kill removal for all entities, no intervention/assistance from IOS operators, etc. Benchmarks terminated under one the following conditions: All F-16s dead, all air adversaries dead, enemy strikers reached their target, or 13 minutes elapsed time. During the course of the study, the vast majority of benchmarks terminated under one of the first three rules.

The participants' overriding goal for the point defense benchmark scenario was to prevent the enemy strikers/bombers from reaching the base – success being striker denial or kill. The second and third most important goals are to minimize friendly mortalities and maximize the adversary kills. The point defense benchmark scenarios were selected for examination in the present study as pre- and post-test assessments because:

(a) point defense scenarios have very clear goals and measures of success,
(b) all the benchmark engagements have equivalent levels of complexity,
(c) three benchmark scenarios occur at the beginning <u>and</u> the end of the week-long DMO syllabus,
(d) the same pilots in the same cockpit assignments perform the mirror-image benchmark scenarios at the beginning and the end of the week (unknown to them), and

(e) the benchmarks were flown under real-time kill removal and strict data collection rules.

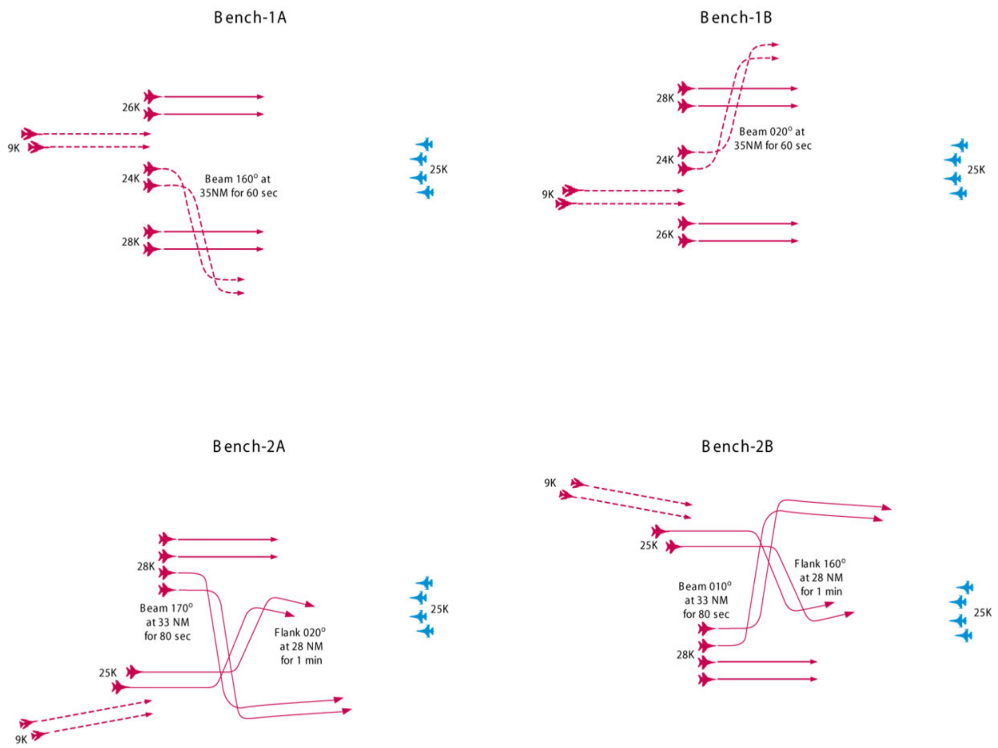

Figure 1 Example mirror-image point defense benchmark scenarios used for the pre- and post-test.

The MEC-based building-block training began immediately after the benchmarks (with the remaining time during session two) on Monday afternoon and continued through the course of the week. Participating teams were exposed to four to eight full engagements per session. While these training sessions emphasized Defensive Counter Air (DCA) scenarios, pilots also flew Offensive Counter Air (OCA) and air-to-ground missions. Usually, participating teams experienced about 35 training engagements between the Monday and Friday benchmarks, providing an intensive training curriculum. The building block training sessions progressed in complexity by increasing the number of threat aircraft, the type of threat aircraft, the threat aircraft reactivity/maneuver, and/or an increase in the vulnerability time.

Either after the last session on Thursday or on Friday morning, pilots took the second Pathfinder exercise and were given a DMO reaction rating form. The DMO rating form is a rating scale survey that pilots use to rate their training experience at DMO. After the last session on Monday and Friday, the team was also given a self-report feedback form with open-ended questions asking if they felt their objectives have been met and what facilitated or hindered their performance. Finally, before departure, teams were given a performance outbrief after their last set of benchmarks. This outbrief consisted of graphs for a number of the objective measures, revealing the team’s performance.

RESULTS

Dataset Class A: Objective Data

Selected outcome and skill metric results from Schreiber, Stock, and Bennett, (2006b) are summarized in Table 2. Out of three scheduled pre- and post-test benchmarks, at least two scenarios must have been successfully flown and recorded to be included in the objective analyses. For those 53 teams, each team's pre-test benchmark performance (on either two or three scenarios) was averaged to produce a benchmark session average. The same was done for post-test benchmark performance. All objective analyses in Table 2 consisted of t-tests. As can be seen in Table 2, there were significant improvements in mission outcome measures and many skill/process measures. Additionally, all significant effects were in the expected direction (i.e., improved performance).

Table 2 Summary results for 17 selected Schreiber, Stock, and Bennett, (2006) objective metrics (NS = Not Significant).

Variable Name	Change from Mon-Fri (%)	p-value
"Top Gun" scoring scheme (composite score of fratricides, strikers killed before or after target, and hostile fighter and F-16 mortalities)	Increased 314.21%	<.01
# of enemy strikers reaching target	Decreased by 58.33%	<.01
Closest distance achieved in #1	Increased by 38.10%	<.04
# of Viper mortalities	Decreased by 54.77%	<.01
# of enemy strikers killed (before reaching base)	Increased by 75.26%	<.01
# of enemy aircraft killed	Increased by 9.20%	<.01
Proportion of Viper AMRAAMs resulting in a kill	Increased by 6.82%	<.03
Proportion of Threat ALAMOs resulting in a kill	Decreased by 51.60%	<.01
Avg time allowing hostiles into MAR (sec)	Decreased by 55.20%	<.01
Avg time allowing hostiles into N-pole (sec)	Decreased by 60.33%	<.01
Slant range at AMRAAM pickle	Increased 10.31%	<.01
Mach at AMRAAM pickle	Increased 5.28%	<.01
Altitude at AMRAAM pickle	Increased 7.97%	<.01
Loft angle at AMRAAM pickle	Increased 14.80%	<.01
G-loading at AMRAAM pickle	NS	NS
F-Pole range (hits and misses)	Increased 8.12%	<.01
A-pole range	Increased 14.35%	<.01
# of communication step-overs (Viper flight)	Decreased 16.33%	<.01

Dataset Class B: SME Observer Rating Data

Summary results for the SME observer real-time and blind rating data is presented in Table 3. Out of three scheduled pre- and post-test benchmarks, at least two scenarios must have been successfully flown and recorded to be included in the subjective analyses. For those 37 teams, each team's pre-test benchmark performance (on either two or three scenarios) was averaged to produce a benchmark session average. The same was done for post-test benchmark performance.

Table 3 Summary results for SME real-time and blind observer rating data.

Construct Rated	Monday real-time			Friday real-time			Monday blind			Friday blind		
	N	M	S.E.	N	M	S.E.	N	M	S.E.	N	M	S.E.
Brief: Mission Prep	29	1.97	0.13	29	2.90	0.11						
Brief: Developing Plan	29	1.69	0.13	29	2.83	0.13						
Brief: Organization	29	1.83	0.15	29	2.93	0.11						
Brief: Content	29	1.59	0.16	29	2.76	0.11						
Brief: Delivery	29	1.83	0.15	29	2.62	0.12						
Brief: Instructional Ability	29	1.62	0.15	29	2.55	0.11						
Brief: Sys Knowledge	29	1.86	0.15	29	2.66	0.10						
Brief: Overall Quality	29	1.55	0.13	29	2.90	0.08						
Radar Mech: El Strobe	48	1.15	0.09	48	2.35	0.13	36	2.06	0.14	36	2.42	0.12
Radar Mech: Range Control	48	1.56	0.09	48	2.54	0.11	36	2.08	0.13	36	2.39	0.11
Radar Mech: Azimuth Control	48	1.58	0.10	48	2.71	0.14	36	2.08	0.13	36	2.44	0.08
Radar Mech: Util.Correct Mode	48	1.44	0.10	48	2.56	0.11	36	2.08	0.15	36	2.42	0.10
Gameplan - Tactics	48	1.81	0.09	48	2.71	0.09	36	2.28	0.11	36	2.61	0.08
Gameplan: Execution	48	1.17	0.10	48	2.42	0.09	36	1.64	0.11	36	2.22	0.11
Gameplan: Adj..on-the-fly	48	0.94	0.10	48	2.31	0.13	36	1.33	0.11	36	2.06	0.12
TI: Formation	48	1.17	0.12	48	2.33	0.12	36	1.86	0.11	36	2.17	0.11
TI: Detection / Commit	48	1.69	0.09	48	2.88	0.11	36	2.25	0.10	36	2.69	0.08
TI: Targeting	48	1.56	0.13	48	2.58	0.12	36	2.22	0.14	36	2.61	0.10
TI: Sorting	48	1.21	0.12	48	2.44	0.12	36	1.83	0.12	36	2.34	0.13
TI: BVR launch and leave	48	1.08	0.11	48	2.21	0.12	36	1.92	0.13	36	2.46	0.11
TI: BVR launch and react	48	1.04	0.10	48	2.25	0.11						
TI: Intercept Geometry	48	1.33	0.10	48	2.21	0.10	36	1.56	0.12	36	1.97	0.07
TI: Low Altitude Intercepts	48	0.98	0.11	48	2.10	0.11						
Engagement Decision	48	1.13	0.12	48	2.23	0.11	36	1.81	0.12	36	2.25	0.11
Spike Awareness	48	1.25	0.10	48	2.52	0.13	36	1.86	0.18	36	2.41	0.11
E/F/N Pole	48	1.06	0.10	48	2.23	0.13	36	1.61	0.15	36	2.00	0.13
Egress / Separation	48	1.02	0.11	48	2.35	0.11	36	1.59	0.15	36	2.22	0.12
AAMD: RMD	48	0.98	0.11	48	2.40	0.12						
AAMD: IRCM	48	1.23	0.10	48	2.40	0.09						
AAMD: Chaff / Flares	48	1.17	0.11	48	2.60	0.09						
Contracts	48	1.13	0.10	48	2.29	0.11	36	1.75	0.11	36	2.31	0.11
ROE Adherence	48	1.27	0.12	48	2.31	0.12	36	2.80	0.20	36	2.91	0.22
ID Adherence	48	1.25	0.14	48	2.35	0.15	36	2.86	0.21	36	2.83	0.22
Post Merge Maneuvering	48	1.25	0.11	48	2.38	0.09	36	1.44	0.10	36	2.11	0.11
Mutual Support	48	0.90	0.10	48	2.23	0.14	36	1.50	0.13	36	2.11	0.12
Visual Lookout	48	1.08	0.08	48	2.21	0.11	36	1.36	0.13	36	2.25	0.12
Weapons Employment	48	1.27	0.11	48	2.50	0.10	36	2.33	0.11	36	2.69	0.10
Clear Avenue of Fire	48	1.73	0.12	48	2.56	0.12						
Comm: 3-1 Comm	48	0.98	0.06	48	2.25	0.10	36	1.64	0.11	36	1.92	0.12
Comm: Radio Discipline	48	1.06	0.08	48	2.29	0.10	36	1.75	0.10	36	2.17	0.12
Comm: GCI Interface	48	1.08	0.09	48	2.63	0.12	36	1.89	0.11	36	2.28	0.11
Fuel Management	48	1.65	0.12	48	2.60	0.13						
Flight Discipline	48	1.17	0.12	48	2.15	0.11	36	1.83	0.09	36	2.19	0.10

	Monday real-time			Friday real-time			Monday blind			Friday blind		
Situation Awareness	48	0.96	0.10	48	2.33	0.12	36	1.44	0.12	36	2.17	0.09
Judgment	48	1.02	0.11	48	2.25	0.12	36	1.50	0.11	36	2.17	0.09
Flight Leadership/Conduct	48	1.08	0.10	48	2.44	0.13	36	1.50	0.13	36	2.36	0.11
Briefed Objectives Fulfilled	48	0.85	0.09	48	2.44	0.12	36	1.47	0.10	36	2.31	0.12
Overall Engagement Grade	48	0.81	0.09	48	2.35	0.12						
Debrief: Organization	29	1.87	0.15	29	2.93	0.12						
Debrief: Reconstruction	29	1.97	0.17	29	2.86	0.12						
Debrief: Delivery	29	1.90	0.18	29	2.86	0.13						
Debrief: Analysis	29	1.84	0.17	29	2.93	0.11						
Debrief: Instr. Ability	29	1.71	0.17	29	2.72	0.11						
Debrief: ID Adherence	29	1.68	0.19	29	2.59	0.14						
Debrief: Flight Leadership	29	1.52	0.15	29	2.69	0.10						
Debrief: Miss Obj's Accomp	29	1.35	0.13	29	2.79	0.12						
Debrief: Overall Quality	29	1.68	0.16	29	2.86	0.10						

***Note: For the blind ratings, briefs/debriefs could not be observed. Other blind rating empty cells reflect missing data.**

For the real-time ratings, there were a total of 57 constructs to be rated, and the average ratings for a given construct ranged from .81-1.97 (mean = 1.36) on Monday to 2.10-2.93 (mean = 2.51) on Friday. The real-time ratings of the brief and debrief were analyzed separately from the engagement data because (a) we felt this was a natural assessment distinction from assessing actual simulator "flying," and (b) during a session, only one brief and one debrief period surrounded an hour of flying multiple SME-evaluated engagements (refer to Table 1). Therefore, there were less assessment data for the brief and debrief than for the engagements. For the brief and debrief data, there were 29 paired Monday and Friday benchmarks with complete data for all 17 brief and debrief constructs. These data showed that the SMEs rated participants significantly higher on Friday's brief and debrief (mean = 2.76) than on Mondays (mean = 1.75), $F(1, 28) = 97.22$, $p < 0.001$. For the engagement portion of the gradesheet, there were 50 pairs of benchmarks for which we had complete data from an SME rater on all 40 "flying" constructs for both Monday and Friday. Over all engagement "flying" constructs, an analysis of variance showed that the gradesheet scores were significantly higher on Friday's benchmarks (mean = 2.40) compared to Monday's benchmarks (mean = 1.20), $F(1, 47) = 150.86$, $p < 0.001$. Follow-up t-tests revealed that for *all* 57 real-time rated constructs, Friday's score was significantly higher than Monday's ($p < 0.001$ for all).

Since all 57 constructs were significantly higher on Friday and the Krusmark et al. (2004) study suggested a possible lack in measurement sensitivity, we performed an exploratory factor analysis on the real-time rating data. Separate principle component analyses were run on the Monday and Friday real-time ratings. Scree plots revealed that just three factors seem to underlie both Monday and Friday ratings. A maximum likelihood procedure was run with a limit of three factors, again separately for Monday and Friday's data. This showed that, of the three factors, there was some overlap in the first two factors, but the third factor was different on Monday and Friday. Additionally, a large number of the constructs had factor loadings above .35.

For the blind ratings, there were 36 matched Monday and Friday benchmarks. For the 32 constructs for which we had sufficient data, average ratings for a given construct ranged from 1.33-2.86 (mean = 1.85) for Monday to 1.92-2.91 (mean = 2.33) for Friday. These differences between the ratings on the Monday benchmarks and the Friday benchmarks were significant, $F(1, 35) = 14.588$, $p = .001$, even though the raters did not know what day's benchmark they were watching. Follow-up t-tests revealed that 27 of the 32 constructs were significant ($p < 0.05$). Two of the constructs (Radar Mechanics – range control and Radar Mechanics – Utilizing correct mode) approached significance ($p = 0.094$ and $p = 0.076$, respectively). Three of the constructs (E/F/N pole, ROE adherence, and ID adherence) were not significant ($p > 0.1$).

Dataset Class C: Participant Survey Data

Criterion for participants' survey data to be included in the dataset was complete 5-day DMO participation and completed survey forms. There were 327 pilots and 49 AWACS operators who completed 58 ratings of DMO and answered six open-ended questions. As reference, the rating scale used was:

1 = Strongly Disagree
2 = Somewhat Disagree
3 = Somewhat Agree
4 = Strongly Agree

Results from the ratings show favorable DMO ratings from both pilots and AWACS operators across the 58 statements. For pilots, the average statement ratings ranged from a low of 2.56 ("As a result of this training, I have improved my VID tactics") to a high of 3.94 for two statements ("I would recommend this training experience to other pilots/controllers" and "DMO will positively impact my combat mission readiness"). For AWACS operators, the average statement ratings ranged from a low of 2.45 ("This training provided excellent experience in radar mechanics") to a very high, almost unanimous score of 3.98 ("I would recommend this training experience to other pilots/controllers"). For the pilots and AWACS participants only 4/58 and 9/58 of the average individual statement ratings, respectively, were below a 3.0. The 58 statements were grouped into seven summary categories and the weighted mean ratings for each summary category are provided in Table 4. As can be seen, all summary category mean ratings were above 3.0.

Since most of the open-ended questions were created to quickly identify any emerging technical issues for the Mesa research site to correct, we performed a simple content analysis on just one of the open-ended questions. Specifically, *"List the top five things you feel were beneficial about the training you received here at DMO. Next to each item, please state why it was beneficial."*

Table 4 Summary categories for the 58 statements participants were asked to rate on the 1-4 scale. The number of statements for each category and the nature of the statements are provided in the parenthetical note.

Seven Summary Categories	**Pilots Weighted Mean (n; s.e.)**	**AWACS Weighted Mean (n; s.e.)**
Overall DMO Training Value (20 statements relating to improving various skills, valuable use of time, positively impacting readiness, etc.)	**3.69 (6521; .03)**	**3.58 (915; .09)**
DMO expectations (6 statements relating to having high expectations and goals being met)	**3.63 (1949; .03)**	**3.59 (279; .10)**
DMO Opinions (11 statements relating to desire to see DMO expanded, recommending DMO to others, should be part of spin-up training, etc.).	**3.72 (3591; .03)**	**3.69 (527; .07)**
Home Unit Conditions (6 statements relating to not being able to get similar training at home unit and degree to which the DMO training will maintain skills used at home unit).	**3.40 (1959; .04)**	**3.01 (275; .13)**
DMO General Statements (4 statements relating to realistic visual scenes, accurate representation of operational missions, accurate databases, and sufficient fidelity).	**3.20 (1304; .04)**	**3.01 (181; .12)**
DMO Scenario Characteristics (3 statements relating to scenario realism and intel relevance).	**3.23 (953; .04)**	**3.23 (137; .10)**
DMO Syllabus Mission Flow (8 statements relating to overall pace/flow of missions, mission organization, appropriate level of difficulty, etc.)	**3.57 (2612; .03)**	**3.53 (378; .09)**

Before calculating item frequencies and percentages, we checked the reliability of the coded segments using 80% agreement as our criterion, and both pilot (91.6%) and AWACS (80.7%) data reached this criterion. The percentage of statements per category are provided for both the pilots and the AWACS participants in Table 5. With a combined total of over half their comments, pilots most liked the realistic qualities or the skill improvement/acquisition. AWACS operators, on the other hand, responded with over 40% of their comments that they most liked the scenarios or the skill improvement gained from the briefs/debriefs.

Table 5 Comments by category.

Category	% of comments (pilots)	% of comments (AWACS)	Frequency (pilots)	Frequency (AWACS)
Realistic Qualities	28.14%	1.07%	316	2
Skill Improvement/Acquisition	24.49%	2.14%	275	4
Briefs/Debriefs (Training specific facilities)	10.95%	8.02%	123	15
Communication	7.30%	8.56%	82	16
Tactics	6.77%	7.49%	76	14
Scenarios (Quantity/Variety/Quality)	3.83%	22.99%	43	43
Controller/AWACS Integration	3.21%	8.02%	36	15
SIM Characteristics	3.21%	7.49%	36	14
Situation Awareness	3.03%	6.99%	34	13
Cold Ops	2.32%	2.67%	26	5
Threats	2.23%	2.14%	25	4
Incidentals (non-DMO references)	1.96%	1.60%	22	3
Other Training Related Benefits	.98%	1.60%	11	3
Weapons/Weapon Employment	.80%	0%	9	0
Briefs/Debriefs (Skill Improvement Acquisition)	.53%	18.72%	6	35
Briefs/Debriefs (Non-Specific)	16%	.53%	3	1

The pilots rated each of the 45 MEC experiences (e.g., “task saturation,” “lost mutual support,” “full range of adversary air threat and mix,” etc.; see Schreiber, Rowe, and Bennett (2006d) for the complete list) on a 5-point scale as to the extent that different environments provide training for that experience. For reference, the scale used was:

0 = Not at all/Does Not Apply
1 = To a Slight Extent
2 = To a Moderate Extent
3 = To a Substantial Extent
4 = To a Great Extent

The MEC experiences by environments survey was administered late in the study and 32 pilots completed the survey. Results are shown in Table 6. Across all 45 MEC experiences, the average ratings per environment were computed and a within-subjects ANOVA performed. The differences in average ratings between environments was found to be highly significant, $F(7,217) = 11.96$, $p < .01$, with the Mesa DMO environment rated highest overall and the Weapons and Tactics Trainer/Desk Top Trainer (WTT/DTT) environment rated lowest overall. Contrast tests comparing the Mesa DMO environment average against the average of all other environments revealed that Mesa was rated significantly higher (at alpha = .01) than all but one other environment (the only exception being the RAP Flag/CFTR environment category). Therefore, it was not surprising to find that the distribution of ratings varied with environment, as shown in Table 11. Ratings for the WTT/DTT and Operation Northern/Southern Watch (ONW/OSW) environments were generally negative, while ratings for Mesa DMO were most positive. Only 3 of the 8 environments (Mesa DMO and the two RAP environment categories) were judged to provide half or more of the MEC experiences “to a moderate extent” or better.

Table 6 Ratings averaged over all 45 MEC experiences

Environment	Avg rating over all 45 experiences	% of experiences rated 3 or higher	% of experiences rated 2 or higher	% of experiences rated 1 or higher
Mesa DMO	2.65	40%	84.4%	95.6%
RAP Flag/CFTR	2.37	4.4%	75.6%	100%
RAP except Flag/CFTR	2.09	4.4%	60%	97.8%
UTD	1.85	0	44.4%	93.3%
Sustained Combat Ops	1.74	2.2%	33.3%	91.1%
MTC/FMT	1.54	0	11.1%	86.7%
ONW/OSW	1.08	0	0	60%
WTT/DTT	0.93	0	0	40%

Individual cell rating results show drastically different averages for each of the 8 environments across the 45 various experiences (i.e., the 360 cells), ranging from the lowest rating of just .25 (tied) for two environments, Mission Training Center/Full Mission Training (MTC/FMT) and WTT/DTT, providing the "G-induced physical limitations" experience, to a high of 3.66 for the Mesa DMO environment providing the "1:3 + Force Ratio" experience. For each of the 45 MEC experiences, the highest and lowest rated environment was identified and tabulated. The Mesa DMO environment was rated as best (or tied as best) for providing 29/45 (64.4%) of the MEC experiences, while the WTT/DTT environment category was rated worst (or tied for worst) for 33/45 (73.3%) of the MEC experiences. For further reporting of the user opinion survey data, we refer the reader to Schreiber, Rowe, and Bennett (2006d).

Dataset Class D: Pathfinder

Using a 9-point scale, 144 F-16 pilots rated the relatedness of 105 pairs of air combat concepts. The concepts used and then paired together for relatedness ratings were:

Flight lead	Wingman	Weapons Director
Sanitize AOR (area of responsibility)	Make threat calls	Clear avenue of fire
Linebacker for leading edge	Listen	Mission flow
Build picture	Allocate radars	WEZ denial
Formation/visual mutual support	Radar work to support shot	Target as assigned

Participants were asked to perform the same paired relatedness judgments both before and after the five days of air combat DMO training. The ratings among concept pairings are assumed to provide an estimate of the distance between concepts in memory. We expected that the Pathfinder tool, at the onset of DMO training, would reveal different knowledge structures as a function of F-16 experience level. As less experienced pilots learn more about the air-to-air combat concepts while flying in DMO exercises, results from post-DMO training concept ratings should reveal that their knowledge structures will become more stable and will reflect the permanence of the more expert knowledge structures.

The data were analyzed by occasion (before/after DMO training), flight qualification level (Instructor, Flight Lead, or Wingman), and F-16 cockpit assignment (Viper 1, Viper 2, Viper 3, or Viper 4). Across the six matrices formed by crossing occasion by qualification level of pilots, 14 links were in common across occasions and level of qualification, and the total number of links varied from a minimum of 16 to a maximum of 18. Similarly, for the eight matrices formed by crossing occasion and Viper position, 15 links were common to all Viper positions across both occasions, and the total number of links varied between a minimum of 16 and a maximum of 18. Further, across all matrices, the concepts Linebacker for the Leading Edge and Clear Avenue of Fire emerged at the center of a stable set of nodes. Overall, the patterns of links were very stable and consistent, regardless of pre/post DMO training occasion or by F-16 experience or flight position.

DISCUSSION

DMO training provides opportunities simply not obtained elsewhere. Obviously, new training techniques and technologies are much more easily assessed and addressed in DMO than with an actual airframe. Increased training in DMO environments is likely to have a number of indirect financial benefits (e.g., OPSTEMPO, or, with distributed capability, a reduction in travel expenses). Also, unlike stand-alone simulators of the past, pilots can now actually exercise higher order skills and teamwork. Though live-fly exercises also provide this, pilots reported training on these higher order skills as infrequent, or a current training “gap” and that DMO could potentially fill that gap. Furthermore, DMO can provide repetition levels simply not possible with live-fly. In the current work, an F-16 team flew, on average, over 40 total engagements, employing several hundred missile shots against hundreds of threats over just eight, non-familiarization mission sessions. A simulator session in the current work was an hour, creating an average of five many-versus-many scenarios per hour. Pilots of just a generation ago might take an entire career to achieve 40 such experiences that, in the current DMO protocol used here, took just eight hours of simulator time. But, of course, that repetition must result in measurable and significant learning—that is, a more competent warfighter.

The converging results from the different dataset classes reported here provide substantial evidence that pilots become more competent within the simulator as a function of DMO training. We anticipated a number of factors in this study would undermine the chances of revealing statistically significant within-simulator DMO learning effects. Some of these include:

(a) an applied research study on an extremely complex and ecologically valid task,
(b) changes to the experimental environment creating additional noise in the data,
(c) missions/tasks that can be only partially controlled, and

(d) a highly experienced, combat-ready participant pool whose performance levels arguably may have been at or approaching asymptote before ever participating in the current study.

Finding highly significant performance differences between the pre- and post-tests in light of those factors strongly suggests that DMO training yields considerable within-simulator warfighter competency improvement.

Enemy strikers reaching base—the most important combat-relevant metric in the current study—were reduced by an incredible 58.33% on Friday ($p < .01$). Furthermore, F-16 mortalities dramatically decreased by 54.77% ($p < .01$). If this learning effect transferred completely to combat, consider (a) the capability of force difference, (b) the friendly force lives saved, or (c) calculating the Air Force's financial implications from the reduced hard asset loss on a single mission, let alone on an entire campaign. Based solely upon the outcome metrics, effective DMO training assuredly exists, at least for within-simulator improvement.

Further buttressing the DMO within-simulator training benefits, the tremendous improvements in outcome metrics were not achieved by pilots negatively altering their risk tolerance. Even though on the post-test benchmarks the F-16 teams routinely denied enemy strikers and easily disposed of threats compared to the pre-test benchmarks, the F-16s teams did so while reducing their vulnerability exposure. The F-16s launched their AMRAAMs at greater ranges and greatly reduced their exposure to hostiles penetrating MAR and N-pole. The significant differences observed on the outcome metrics were not attributable to a risk/reward trade-off, illustrating skill proficiency gains, especially for weapons employment and controls intercept geometry, two MEC skills specifically evaluated objectively in the current work. Other related signal detection research has also shown that the pilots as a team make more accurate and timely decisions on when to shoot AMRAAMs (Stock, Schreiber, Symons, Portrey, & Bennett, 2004).

Unsurprisingly, the other data sources only reinforce the objective results. The SME observer ratings—both real-time and blind—corroborated the objective results. SMEs rated performance at the end of the week significantly higher than performance at the beginning of the week. Additionally, pilots in general gave favorable ratings on all 58 statements we asked them to rate. Convincingly, pilots highly recommend this training to their peers, where all but 16 of 327 pilots rated the statement, "I would recommend this training experience to other pilots/controllers" with the highest rating possible, "Strongly Agree." Furthermore, when analyzing the pilots' open-ended responses to what they liked best about training in the DMO environment, the most frequent category of comment was skill acquisition. The Pathfinder results, though not revealing a significant change in understanding over the course of the week, do suggest that the pilots' knowledge structures are stable for the abstract concepts used. We speculate that more detailed, training-specific concepts would reveal differences in knowledge structures as a function of DMO training.

General opinion in the DMO community has been that DMO training provides value. The converging results from the different data classes reported here unquestionably support that assertion, perhaps even provide justification for raising expectations of DMO's training potential. However, the current work does not specifically address some potential negative training associated with DMO, such as lack of consequences for running out of fuel and not experiencing any g-force, emergency procedures, or inclement weather during the missions. Additionally, this study was performed on a non-random sample of F-16 pilots on point defense benchmarks, limiting generalizability. Subsequent studies should investigate the effectiveness of random samples in different missions and different domains. Furthermore, many of the results in the current work cannot be delineated between other factors, such as how much of the improvement was due to individual learning by each pilot versus pilots learning to coordinate with one another (i.e., team cohesion)? Or, how much of the learning can be attributable to actual simulator flying versus debriefing? Learning effects in the current work reflect the week's experience in total, and as such, the learning effect cannot be delineated to answer finer detailed questions. Other, very important application-oriented studies must also be undertaken, such as what is the degree of transfer to a live-fly training event? How quickly do these skills decay? The powerful results reported here certainly support a position of aggressively pursuing DMO's training potential, but additional research will help us better understand the benefits and how best to implement DMO training.

REFERENCES

Alliger, G. M., Colegrove, C. M., & Bennett, W. Jr. (2003, April). *Mission Essential Competencies: New method for defining operational readiness and training requirements.* Paper presented at the Thirteenth International Occupational Analyst Workshop. San Antonio, TX.

Alliger, G. M., Garrity, M. J., McCall, J. M., Beer, L., & Rodriguez, D. (2003, April). *Competency-based definition of work and performance for command and control.* Paper presented at the Thirteenth International Occupational Analyst Workshop. San Antonio, TX.

Alliger, G. M., Garrity, M. J., See, K., McCall, J. M., & Tossell C. (2004). Advanced training for commanders: A competency-based approach to training requirements definition for the JFACC. In *2004 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Alliger, G. M., McCall, J. M., Garrity, M. J., Morley, R. M., Beer, L., & Rodriguez, D. (2003). Mission Essential Competencies for the AOC: A basis for training needs analysis and performance improvement. In *2003 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Bell, H. H, & Waag, W. (1998). Evaluating the effectiveness of flight simulators for training combat skills: A review. *The International Journal of Aviation Psychology, 8(3), 223-242*.

Bennett, W. Jr., Schreiber, B. T., & Andrews, D. H. (2002). Developing competency-based methods for near-real-time air combat problem solving assessment. *Computers in Human Behavior, 18(6),*. 773-782.

Berger, S. & Crane, P. M. (1993). Multiplayer simulator based training for air combat. In *Proceedings of 15th Industry/Interservice Training Systems Conference*, Orlando, FL: National Security Industrial Association.

Bills, C. G. & Devol, D. (2003). How F-16 MTC team achieved operational training capability. In *2003 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Boyne, W. J. (2000). Red Flag. *Air Force Magazine*, *83(11)*. Washington, D.C.: Air Force Association. Retrieved Mar. 7, 2005 from http://www.afa.org/magazine/nov2000/.

Brecke, F. H., & Miller, D. C. (1991). *Aircrew performance measurement in the air combat maneuvering domain: A critical review of the literature.* (AL-TR-1991-0042 /

ADB158404). Williams Air Force Base, AZ: Armstrong Laboratory, Human Resources Directorate, Aircrew Training Research Division.

Callander, B. D. (1999). Training in networks. *Air Force Magazine, 82(8).* Washington, D.C.: Air Force Association. Retrieved Mar. 7, 2005 from http://www.afa.org/magazine/aug1999/.

Chapin, M. (2004). Message from the Air Force executive. *I/ITSEC Newsletter, 3(4).* Orlando, FL: National Security Industrial Association. Retrieved Mar. 7, 2005 from http://www.iitsec.org/documents/Nov04_IITSEC.pdf.

Chapman, B., Colegrove, C. M., & Greschke, D. (in press). Transforming operational training in the Combat Air Forces. *Journal of Military Psychology*.

Colegrove, C. M. & Alliger, G. M. (2002, April). *Mission Essential Competencies: Defining combat mission readiness in a novel way*. Paper presented at the NATO RTO Studies, Analysis and Simulation (SAS) Panel Symposium. Brussels, Belgium.

Crane, P. M., Robbins, R., & Bennett, W. Jr. (2000). Using Distributed Mission Training to augment flight lead upgrade training (AFRL-HE-AZ-TR-2000-0111, ADA394919). Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division.

Crane, P. M., Schiflett, S. G., & Oser, R. L. (2000). *Roadrunner '98: Training effectiveness in a Distributed Mission Training exercise* (AFRL-HE-AZ-TR-2000-0026, AD A391423). Mesa, AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division.

Denning, T., Bennett, W. Jr., & Crane, P. M. (2002). Mission complexity scoring in Distributed Mission Training. In *2002 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Denning, T., Bennett, W. Jr., Crane, P. M., & Bell, J. A. (2002). Mission characteristics and complexity in Distributed Mission Training. In *2002 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Fischer, J., Case, B., & Bertin, R. (Eds) (2001). *Guidance, rationale, and interoperability manual for the Real-Time Platform Reference Federation Object Model (RPR FOM). Version 2.0D9v2.* (Draft) Orlando, FL: Simulation Interoperability Standards Organization.

Gehr, S. E., Schreiber, B. T., & Bennett, W. Jr. (2004) Within-simulator training effectiveness evaluation. In *2004 Interservice/Industry Training, Simulation and Education*

Conference (I/ITSEC) Proceedings. Orlando, FL: National Security Industrial Association.

Gray, T. H., Chun, E. K., Warner, H. D., & Eubanks, J. L. (1981). *Advanced flight simulator: Utilization in A-10 conversion and air-to-surface attack training* (AFHRL-TR-80-20, AD A094608). Williams Air Force Base, AZ: Air Force Human Resources Laboratory, Operations Training Research Division.

Gray, T. H., & Fuller, R. R. (1977). *Effects of simulator training and platform motion on air-to-surface weapons delivery training* (AFHRL-TR-77-29, AD A043 648). Williams Air Force Base, AZ: Air Force Human Resources Laboratory, Operations Training Research Division.

Gray, T. H., Edwards, B. J., & Andrews, D. H. (1993). *A survey of F-16 squadron-level pilot training in PACAF* (AL-TR-1993-0041, AD A265053). Williams AFB, AZ: Armstrong Laboratory, Aircrew Training Research Division.

Hagin, W. V., Dural, E., & Prophet, W. W. (1979). *Transfer of training effectiveness evaluation: US Navy device 2B35* (Seville Research Corporation Rep. No. TR 79-06). Pensacola, FL: Chief of Naval Education and Training.

Hagin, W. V., Osborne, S. R., Hockenberger, R. L., Smith, J. P. & Gray, T. H. (1982). *Operational test and evaluation handbook for aircrew devices: Operational effectiveness evaluation* (Rep. No. AFHRL-TR-81-44[II], AD A112 570). Williams AFB, AZ: Air Force Human Resources Laboratory, Operations Training Division.

Houck, M. R., Thomas, G. S., & Bell, H. H. (1991). *Training evaluation of the F-15 advanced air combat simulation* (Rep. No. AL-TP-1991-0047, AD A241 675). Williams AFB, AZ: Armstrong Laboratory, Aircrew Training Research Division.

Hughes, R., Graham, D., Brooks, R., Sheen, R. & Dickens, T. (1982). Tactical ground attack: On the transfer of training from flight simulator to operational Red Flag range exercise. In *Proceedings of the 4th Interservice/Industry Training Equipment Conference*. Washington D.C.: National Security Industrial Association, 16-18 November 1982.

Institute of Electrical and Electronics Engineers (1995). *IEEE Standard for Distributed Interactive Simulation - Application Protocols.* . NY: IEEE Publishing.

Jenkins, D.H. (1982). *Simulation training effectiveness evaluation* (TAC Project No. 79Y-OO1F). Nellis AFB, NV: Tactical Fighter Weapons Center.

Kellogg, R., Prather, E., & Castore, C. (1980). Simulated A-10 combat environment. In *Proceedings of the Human Factors Society 24th Annual Meeting* (pp. 573-577). Santa Monica, CA: Human Factors Society.

Kelly, M.J. (1988). Performance measurement during simulated air-to-air combat. *Human Factors, 30(4),* 495-506.

Kirkpatrick, D. (1975). *Techniques for evaluating training programs*. Alexandria, VA: ASTD.

Krusmark, M., Schreiber, B. T., & Bennett, W. Jr. (2004). *The effectiveness of a traditional gradesheet for measuring air combat team performance in simulated Distributed Mission Operations*. (AFRL-HE-AZ-TR-2004-0090). Air Force Research Laboratory, AZ: Warfighter Readiness Research Division.

Lacy, L. & Tuttle, C. (1994). Use of the set data PDU and the event report PDU to support data analysis and reporting. In *Proceedings of the 10th Workshop on Standards for the Interoperability of Defense Simulations*, March 14-18, 1994.

Leeds, J., Raspotnik, W. B., & Gular, S. (1990). *The training effectiveness of the simulator for air-to-air combat* (Contract No. F33615-86-C-0012). San Diego, CA: Logicon.

Lintern, G., Sheppard, D., Parker, D. L., Yates, K. E., & Nolan, M. D. (1989). Simulator design and instructional features for air-to-ground attack: A transfer study. *Human Factors, 31*, 87-100.

McGuiness, J., Bouwman, J. H., & Puig, J. A. (1982). Effectiveness evaluation for air combat training. In *Proceedings of the 4th Interservice/Industry Training Equipment Conference* (pp. 391-396). Washington D.C.: National Security Industrial Association, 16-18 November 1982.

McLean, T. & Roberts, D. W. (1995). Data collection and engineering analysis in the DIS environment: Proposed changes to data, data query, and acknowledge PDUs (ADD210936). In *Proceedings of the 12th Workshop on Standards for the Interoperability of Distributed Simulation*, Orlando, FL, March 13-17, 1995.

Payne, T. A. (1982). *Conducting studies of transfer of learning: A practical guide.* (AFHRL-TR-81-25, AD A110569). Williams Air Force Base, AZ: Air Force Human Resources Laboratory, Operations Training Division.

Payne, T. A., Hirsch, D. L., Semple, C. A., Farmer, J.R., Spring, W. G., Sanders, M. S., Wimer, C. A., Carter, V. E., & Hu, A. (1976). *Experiments to evaluate advanced flight simulation in air combat pilot training: Vol 1. Transfer of learning experiment.* Huntsville, AL: Northrop Corporation.

Robinson, J. C., Eubanks, J. L., & Eddowes, E. E. (1981). *Evaluation of pilot air combat maneuvering performance changes during TAC ACES training.* Nellis Air Force Base, NV: U.S. Air Force Tactical Fighter Weapons Center.

Schreiber, B. T., Stock W. A., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Metric development and* *objectively quantifying the degree of learning.* (AFRL-HE-AZ-TR-2006-0015-Vol II). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Schreiber, B. T., DiSalvo, P, Stock, W. A., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Using the Pathfinder methodology to assess pilot knowledge structure changes*. (AFRL-HE-AZ-TR-2006-0015-Vol V). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Schreiber, B. T., Gehr, S. E., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Real-time and blind expert subjective assessments of learning.* (AFRL-HE-AZ-TR-2006-0015-Vol III). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Schreiber, B. T., Rowe, L. J., & Bennett, W. Jr. (2006). *Distributed Mission Operations within-simulator training effectiveness baseline study: Participant utility and effectiveness opinions and ratings.* (AFRL-HE-AZ-TR-2006-0015-Vol IV). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

Schreiber, B. T., Watz, E. A., & Bennett, W. Jr. (2003). Objective human performance measurement in a distributed environment: Tomorrow's needs. In *2003 Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings*. Orlando, FL: National Security Industrial Association.

Schreiber, B. T., Watz, E., Bennett, W. Jr., & Portrey, A. (2003). Development of a Distributed Mission Training automated performance tracking system. In *Proceedings of the Behavioral Representations in Modeling and Simulation (BRIMS) Conference*. Scottsdale, AZ.

Stock, W., Schreiber, B. T**.**, Symons, S., Portrey, A., & Bennett, W. Jr. (2004, November). *Can performance of air combat teams in Distributed Mission Operations training exercises be described by signal detection parameters?* Paper presented at the annual meeting of the Society for Judgment and Decision Making. Minneapolis, MN.

Symons, S., France, M., Bell, J., & Bennett, W. Jr. (2006). *Linking knowledge and skills to Mission Essential Competency-based syllabus development for Distributed Mission Operations* (AFRL-HE-AZ-TR-2006-0041). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division.

United States Dept. of the Air Force: Flying Operations. (2002, August). *Air Force Instruction (AFI) 11-2, F-16, Volume 1*. Retrieved February 16, 2005, from http://www.e-publishing.af.mil/pubfiles/af/11/afi11-2f-16v1/afi11-2f-16v.

Waag, W. L. (1991). The value of air combat simulation: Strong opinions but little evidence. In *Training Transfer - Can We Trust Flight Simulation? Proceedings of the Royal Aeronautical Society Conference on Flight Simulation and Training.* London: Royal Aeronautical Society.

Waag, W. L., Houck, M. R., Greschke, D. A., & Raspotnik, W. B. (1995). Use of multiship simulation as a tool for measuring and training situation awareness. In AGARD Conference Proceedings 575, *Situation Awareness: Limitations and Enhancement in the Aviation Environment* (AGARD-CP-575). Neuilly-Sur-Seine, France: Advisory Group for Aerospace Research & Development.

Watz, E. A., Keck, L., & Schreiber, B. T. (2004). Using PETS software to capture complex objective measurement data from Distributed Mission Operations (DMO) environments. In *Proceedings of the 2004 Spring Simulation Interoperability Workshop*. Arlington, VA. Paper available as 04S-SIW-143.

Watz, E. A., Schreiber, B. T., Keck, L., McCall, J. M., & Bennett, W. Jr. (2003). Performance measurement challenges in Distributed Mission Operations environments. [Electronic Version] *Simulation Technology Newsletter.* 03F-SIW-022. March 16, 2004. Retrieved 7/26/04 from http://www.sisostds.org/webletter/siso/Iss_102/index.htm 2003 Fall SIWzie Award winner - Fall Simulation Interoperability Workshop.

Wiekhorst, L. A., & Killion, T. H. (1986). *Transfer of electronic combat skills from a flight simulator to the aircraft* (AFHRL-TR-86-45, AD C040549). Williams Air Force Base, AZ: Air Force Human Resources Laboratory, Operations Training Division.

ACRONYMS

AAA	Antiaircraft artillery
AFRL/HEA	Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division
AIM	Air Intercept Missile
ALQ	It designates a countermeasure system on a tactical aircraft.
AMRAAM	Advanced Medium Range Air-to-Air Missile
ANOVA	Analysis of Variance
ATES	Automated Threat Engagement System
AWACS	Airborne Warning and Control System
DCA	Defense Counter Air
DIS	Distributed Interactive Simulation
DMO	Distributed Mission Operations
ECM/ECCM	Electronic Countermeasures/ Electronic Counter Countermeasures
ID	Identification
IEEE	Institute of Electrical and Electronics Engineers
IOS	Instructor Operator Station
MAR	Minimum Abort Range
MEC	Mission Essential Competencies
MTC/FMT	Mission Training Center/Full Mission Trainer
OCA	Offensive Counter Air
OPSTEMPO	Operations Tempo
RAP	Ready Aircrew Program
ROE	Rules of Engagement
ROI	Return on Investment
SADL	Situation Awareness DataLink
SAM	Surface-to-Air Missile
SME	Subject Matter Expert
TEW	Threat Early Warning
USAF	United States Air Force
VID	Virtual Image Display
WTT/DTT	Weapon and Tactics Trainer/Desk Top Trainer